Informatics Bulletin, Faculty of Computers and Artificial Intelligence, Helwan University
Published Online ***Vol 3 Issue 2, July 2021***
(https://fcihib.journals.ekb.eg)

Automated Sports Player Identification: A Survey

Ahmed Nady∗, Ayman Atia† ‡, Amal Elsayed Aboutabl∗
∗ Computer Science Department, Faculty of Computers and Information, Helwan University.
† HCI-LAB, Department of Computer Science, Faculty of Computers and Artificial Intelligence, Helwan University, Helwan, Egypt.
†Faculty of Computer Science, October University for Modern Sciences and Arts (MSA), Cairo, Egypt.
{ahmednady, ayman, amal.aboutabl}@fci.helwan.edu.eg

***Abstract*—Sports can not be said it is a mean for social cohesion by allowing people interaction disregarding their social status, age, etc. only but also it became the entertainment industry from governmental views. Consequently, a large amount of resources has been directed to sports to improve understanding, performance and presentation. Sports video analysis is a popular tool to capture low-level and high-level analysis. In order to make the automatic sports video analysis realized, there is a necessity for player identification across sports. Identifying players in sports videos is a major research challenge due to low video resolution per player, camera perspective, variation of the posture of player, illumination conditions, variety of sports fields and t-shirts. Here, we survey the most significant methods and problem formulations that have been proposed to address the problem of player identification.**

***Index Terms*—Sports Video Analysis, Jersey Number Recognition, Player Identification.**

I. Introduction

Recently, sports video analysis has drawn attention especially in team sports such as soccer, ice hockey, basketball, and volleyball caused by the growing demand from sports professionals and fans for the extraction of semantic information. Results of sports analysis can be utilized for many applications such as storytelling on TV, player statistics such as distance covered, number of shots, etc., evaluation of strengths or weaknesses of a team, tactical analysis, and player performance assessment. Currently, the annotation of sports games performed manually or semi-automatically to identify players and to log every event (e.g., pass, shot, foul) that takes place during the game from video streams [1]. The manual process is cumbersome, error-prone and time-intensive. There is a necessity for an automated sports video analysis system for identification and tracking of players and detection and tagging of events in the game. The steps of team sports video analysis are detection of ball and players in each frame, their tracking over time and analyzing the interaction between ball and players. Sports players tracking is considered a challenging task because of (1) The similar appearance of the players within the team, (2) Occlusion, (3) Unpredicted movement patterns of players as shown in Figure 1. In the tracking phase, a tracker can create a track identity for a player that may be disappeared due to exiting from the camera viewpoint or occlusion for multiple consecutive frames. Hence,

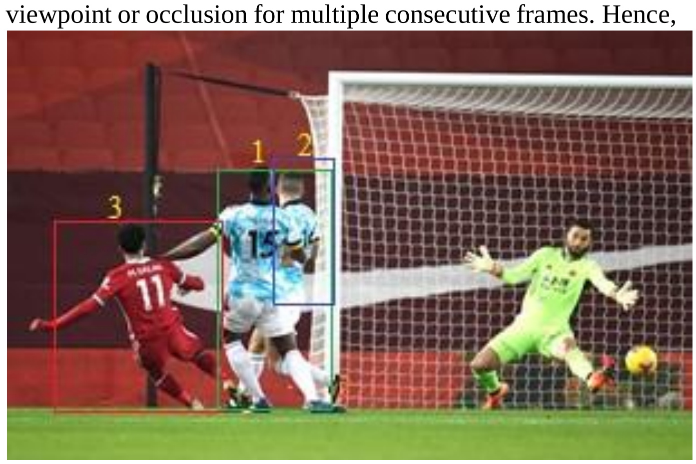

Fig. 1: Tracking multiple players' challenges (Picture from https://images.app.goo.gl/NrrTbmvRV2GWZm1u6).

a new track identity can be created once the player entering camera viewpoint or occlusion resolving. Also, the identities of trackers can be switched due to complex interaction and occlusion. Hence, player identification is the main research challenge for achieving the automatic sports analysis merits. Player identification involves binding the player's identity to each track and connecting it with his actions and statistics.

Identifying players in unconstrained sports video has numerous problems regarding low video resolution, camera perspective, players tilting, illumination conditions, variety of sports fields and t-shirts. Sample frames for these problems are shown in Figure. 2. The features that can be used to identify the player on the court are the face, jersey numbers and tactical positions of players. The player identification approaches can be grouped into four classes: face recognition, jersey number recognition, the tactical position of player and person Re-Identification (see Figure.3). Approaches based on Face recognition for identification work in close-up shots where the player's face is visible and unworkable in overview shots [3], [4]. The jersey number is the visible cue that is common across sports. Owing to t-shirt numbers take a wide portion of the player's back uniform and increasing of HD sports videos, approaches based on jersey numbers recognition are promising [5], [6]. The summary of player identification researches in means of the number of player images in the used dataset, methodology, Sport Genres, input size and accuracy on the corresponding dataset is listed in Table I. From Table I, it is clear that several works used jersey numbers for player identification whether in a specific sport or several sports. For the dataset, there is no benchmark dataset for jersey numbers to compare works of player identification through jersey numbers fairly.

The rest of the paper is organized as follows. Section II reviews jersey numbers recognition based methods. Section III reviews techniques based on person re-identification (person ReId). Section IV reviews Methods based player's tactical position. Section V presents Benchmark datasets. The challenges and future directions are presented in Section VI followed by a summary and conclusion in Section VII.

II. PLAYER IDENTIFICATION BASED JERSEY NUMBER RECOGNITION

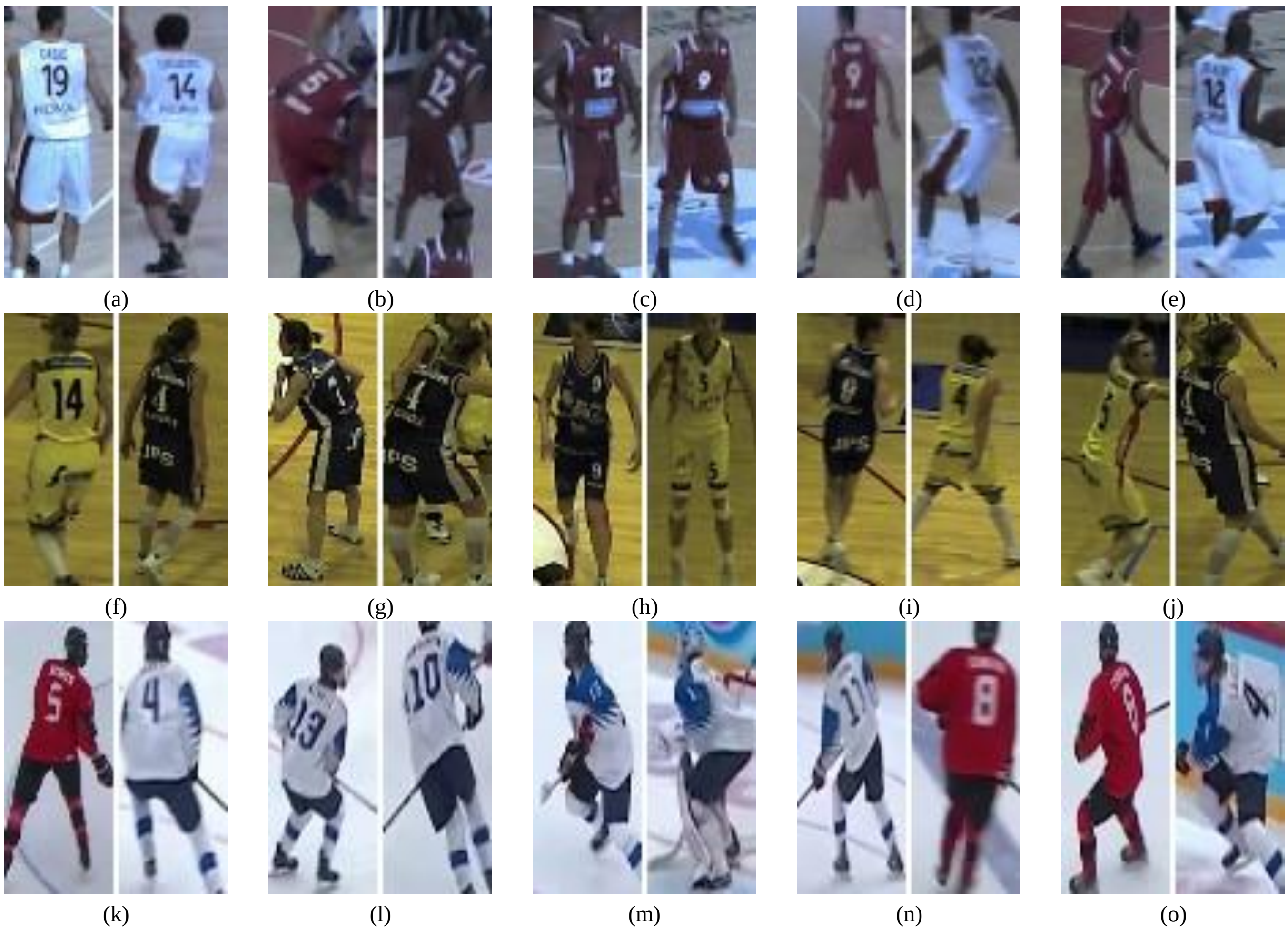

Fig. 2: Illustration of Player identification challenges. Sample Images in each row represent detected players from basketball sport, second basketball sport and ice hockey sport respectively. The players can be detected in various situations: (a) (f) (k) indicate normal situations, (b) (g) (l) pose tilt, (c) (h) (m) Non back jersey numbers, (d) (i) (n) motion blur and (e) (j) (o) severe views. Image obtained from [2].

Jersey number recognition methods can be classified into two groups: traditional approaches and deep learning approaches.

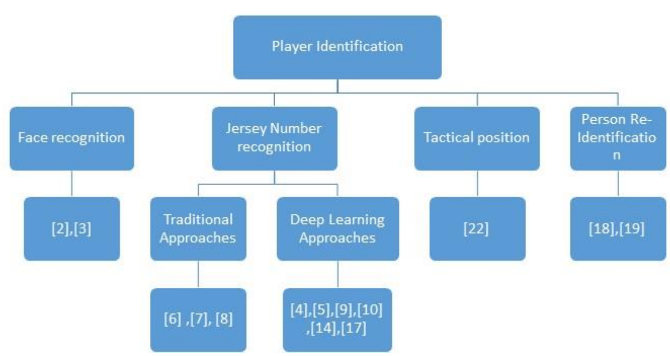

Fig. 3: Player identification approaches and their popular works

A. Traditional approaches

Messelodi et. al. [10] utilize prior knowledge about the color of text background to localize an athlete's name or number on its Bib then textual content of candidate regions are recognized through OCR system. In [11], Lu et. al. localizes the region of jersey number from the bounding box of basketball players using gradient difference, and then OCR scheme is adapted for recognition. Sari et. al. [12] make the OCR module comeˇ after detecting the region of jersey number in HSV color space depending on internal contours. Previous OCR-based researches have limitations of applicability in wide-ranging situations due to the adaptation of handcrafted features.

B. Deep learning approaches

Gerke et al. [5] crop the upper part of the player bounding box from overview shots in broadcast soccer videos. After that, they developed convolutional neural network architecture consisting of three convolutional layers and three fully connected layers for classifying the cropped region. Their finding showed that the accuracy of jersey number recognition improved significantly compared to preceding works [10], [11], [12]. The accuracy of the holistic number approach where every number is treated as a separate class is better than the accuracy of the digit-wise approach in which each digit is treated as a class. Li et al. [6] integrate their CNN architecture with spatial transformer network (STN) to bring attention and affine transformation for the region of jersey number in the bounding box of a soccer player. There is no necessity to cut the upper part of the player bounding box as in [5] to input to the CNN model because STN is utilized for his purpose.

The digit-wise approach [5], [6] struggles to separate the digits of jersey number and variation of camera viewpoints may introduce more difficulty. Liu et al [7] presented a joint framework depending on faster R-CNN to detect sports players and recognize jersey numbers. they addressed the challenges of jersey numbers such as changes of perspective and player postures by means of key points of a player's body that are predicted by a pose-guided regressor. They proposed Region Proposal Network (RPN) which generates bounding boxes for two classes: player and digit and then link person-and-number proposals while keeping only the digit's proposal that is found in the personal proposal. They gathered their dataset with pan and zoom which makes the field of view of the camera narrow and thus jersey numbers become clear, which is not the situation in a broadcast soccer video. Their attempt to generalize across sports demonstrated good localization for players and digits but the performance of digit classification decreased resulting from the different font size.

TABLE I: Comparison of related works for player identification.

Work	Number of player images (Dataset)		Input size	Methodology	Sport Genres	Accuracy	
[5]	8281		40^2	Jersey number classification using CNN	Soccer	83%	
[6]	215036		200^2	Jersey number classification using CNN fused with STN	Soccer	86.7%	
[7]	3567		512^2	Pose-Guided R-CNN for players detection and jersey numbers recognition	Soccer	92.14%	
[8]	166,629		—	Player identification based Re-Id approach using handcrafted features	Basketball	1st Team	61.86%
						2st Team	60.85%
[9]	12500		–	Player identification based Re-Id approach using CNN	Basketball	96%	
[2]	1st Subset	1872	192	Detection of jersey number using fine-tuned CRAFT detector then recognize it through attention-based text recognition model	Basketball and Ice hockey	1st Subset	95.41%
	2^{st} st Subset	851				2^{st} st Subset	85.28%
	3 Subset	1317				3 Subset	85.86%

Nag et al. [13] proposed a framework to detect both jerseys Bib numbers and text on runner's clothes. They used the Single Shot Multibox Detector (SSD) [14] to detect athletes from a given image then employed human pose estimation [15] to extract athlete body's regions that probable containing Bib number that can be affixed to the athlete's clothes in Marathon events or text. After that, the detection of jersey Bib number and text performed on the athlete's torso, left thigh and right thigh separately using the modified pixellink text detection method [16]. Finally, the detected text regions are passed to Convolutional Recurrent Neural Network (CRNN) for recognizing textual content in these regions. In the same manner, Wang et al. [17] tackled the problem of marathon athlete number recognition. They posed the problem as a scene text reading problem. Their model is begun by detecting the athletes using You Only Look Once (YOLOv3) [18] then Connectionist Text Proposal Network (CTPN) [19] text detector is used for detecting athlete's number regions. Finally, the fine-tuned CRNN is utilized to recognize athlete number regions and the interference resulting from text detection can be avoided using the presented tree filtering algorithm.

Nag et al. [13] and Wang et al. [17] used scene text detection and recognition to detect and recognize runners Bib numbers. The bib number is easy to detect due to its horizontal orientation, slight changes in font sizes, and the distinct appearance resulting from having the number on a pure colored background. So, the efficiency of these methods cannot be good enough to recognize the jersey numbers.

Ahmed and Elsayed [2] proposed a framework based on scene text detection and recognition for identifying players through jersey numbers across matches and sports. Their framework comprises three parts. In the first part, the sports players are detected using Yolov4 [20]. In the second part, Character Region Awareness for Text Detection (CRAFT) [21] is used to detect the jersey number region in the player bounding box. Accounting for the challenges of jersey numbers such as player tilting and variation of camera viewpoint, sports fields and jersey numbers, they performed transfer learning via fine-tuning for the text detector on the introduced dataset. In the third part, the textual content of the candidate regions is recognized through the scene text recognition model [22]. Experimental results showed that the proposed framework achieved state-of-the-art performance in different sports.

Chan et al. [23] addressed the player identification problem in ice hockey broadcast videos by utilizing the temporal cue. In their work, they formulated the problem of player identification as a video classification problem where the players can be observed for a while, and thus their jersey numbers can be visible in at least one of the player image sequences. They integrated a residual network (ResNet) with a long short-term memory (LSTM) layer in an end-to-end trainable network to learn the Spatio-temporal visual features of the region of interest of the player image which is the jersey numbers over time. Also, Chan et al. [23] introduced a new way for fusing the output of ResNet+LSTM network that depends on 1-D CNN classifier instead of majority voting manner. The experimental results proved the effectiveness of their presented manner. The limitation of their approach listed as follows: 1) They manually correct the labeling of each frame within the tracklet, So, the multi-player tracker issue (identity switch) due to overlapping or occlusion has not taken into consideration during dataset generation, 2) Their method limited by the collected number of classes,3) Even though the jersey number occupies a small region on the player's t-shirt, they used the entire bounding box of a player instead of focusing on jersey number region to learn features of jersey numbers.

III. Player identification based person Re-Id

The player identification from medium distance in basketball broadcast videos is formulated as a person re-identification problem in which the player is recognized from the whole body [8], [9]. Lu et al. [8] represented the player appearance using a mixture of maximally stable extremal regions (MSER) [24], SIFT features [25], and color histogram features and then they employed a logistic regression classifier for classification. In [9] the player presentation is modeled by fusing the holistic representation from a multi-scale image of the whole player and local salient representations from player parts. Methods of Player Re-Id are not scalable across games and sports due to not existence of all players in the training dataset. Moreover, the t-shirt color should not be changed in all matches and this has difficulty to achieve.

IV. Player identification based player's tactical position

Based on the fact that the jersey number is not always visible due to occlusion, motion blur, unsuitable player pose and the camera viewpoint as well, each player over the pitch has a tactical role such as central defender, winger, forward, etc. and thus their movements should not be randomly but according to its role. Gerke et al.[26] describe the player position by spatial constellations features. The spatial constellation is modeled as a histogram over relative positions of all the team's players. The use of constellation features may increase identification accuracy but cannot be used solely as a single identification mean. So, Gerke et al. [26] proposed a method that combines the spatial constellations features and jersey number recognition. The author poses player identification as an assignment problem given players' trajectories. A player's trajectory represents observations of the same player for certain periods. The use of the assignment problem lets constellation features and jersey numbers be merged flexibly through related cost matrices fusion. The identification accuracy of the method that combines both modalities is gained 13% compared with jersey number recognition only on the presented video clips dataset. The length of each video clip is ranged from 12 to 62 seconds. The method limitation that results from using spatial constellations features is identifying soccer players in each team separately and is not be scalable to other games involving players that are not included in the training set. Finally, the player identification approach is based on manually created player trajectories not tracklets (trajectory fragments) from automatic player tracking

V. Benchmark datasets

There is no publicly available dataset for player identification through each feature used for this purpose. The following datasets: APIDIS and SPIROUDOME Basketball Dataset can be employed to generate a dataset for the player identification where it provides partial manual annotations. In APIDIS Basketball Dataset [27], seven cameras are placed around and above a basketball court and the length of captured video is 16 minutes with 22 frames per second and video resolution is 1600 * 1200. The provided annotations are basketball events that occurred throughout the game and the position for the following entities: player, referee, basketball and ball for one minute. Figure 4. shows the seven views of the dataset.

SPIROUDOME basketball dataset provides 10 minutes basketball video sequence captured by eight stationary cameras positioned around the basketball court. The videos are captured with 25 frames per second and 1600 * 1200 resolution. The annotation is provided only for the player position in camera 2. Figure 5. shows the eight views of the dataset.

VI. Challenges and Future Directions

In recent years, player identification has got interest due to unprecedented success of detection and recognition in computer vision. In this section, we explore the potential directions based on the identified challenges in the literature. The potential directions are listed as follows:

1) Player identification across sports

The past studies presented a player identification system for each sport separately not across sports and some of them even cannot be scalable for other matches from the same sport [6], [7], [9]. The challenges in different sports are listed as follows: 1) jersey numbers have characteristics in each sport such as jersey numbers in soccer and basketball sport are plain number whereas they are bulky with sharp contours in ice hockey and American football, 2) the wide player pose variability that makes the jersey number difficult to be detected and recognized, 3) player image resolution differs per sport due to sports playfield width. For example, soccer player images have low resolution compared to basketball player images. Consequently, there is a necessary demand for a player identification system that performs not only within a single sport but also across different sports by addressing the mentioned challenges.

Fig. 4: Sample images from seven views of APIDIS basketball dataset

Fig. 5: Sample images from eight views of SPIROUDOME basketball dataset

2) Synthetic data for training

To build a reliable player identification system, a huge amount of data (player images) should be gathered and annotated such as Li et al. [6] collect data from 164 soccer matches to identify players through jersey numbers, and still, there is room for improvement. The annotations can be done at the player image level or number/ digit level. The image level annotation means that players' images are annotated with their identity without localizing the visual part responsible for getting this identity. The number/digit level annotation means that players' images are annotated with bounding boxes of jersey numbers/digits and their values. There is no publicly available dataset for player identification and both annotations require a lot of time and effort. Thus, utilizing synthetics approaches can make this process easy to achieve.

3) Employing video information

Exploiting temporal cue can have a high influence on player identification accuracy [8], [23]. Based on the fact that players are observed over a while, the player identification is made per partial tracklet, not per frame. Easy-to-classify player images can be propagated to other images in the player's partial trajectory and hence the player identification performance can be improved.

4) Long-term player identification

Multi-object tracking solutions that are used to associate detections in consecutive frames produce short-term trajectories due to challenges encountered in team sports. To perform fine-grained video analysis for team sports matches, player identification throughout the match is needed. One approach for that is associating partial trajectories that belonging to the same player.

VII. Conclusion

Through this paper, we performed a detailed survey of current techniques on player identification. First, We explored the challenges of the problem of player identification and the importance of having reliable player identification on the whole process of sports video analysis. Second, we divided the player identification techniques according to the used features, and then relevant studies were reviewed. Finally, we reported the potential directions based on the identified challenges in the literature.

References

[1] M. Schlipsing, J. Salmen, M. Tschentscher, and C. Igel, "Adaptive pattern recognition in real-time video-based soccer analysis," *Journal of Real-Time Image Processing*, vol. 13, no. 2, pp. 345–361, 2017.

[2] A. Nady and E. E. Hemayed, "Player identification in different sports," in *International Conference on Computer Vision Theory and Applications*. SciTePress, In Press.

[3] L. Ballan, M. Bertini, A. Del Bimbo, and W. Nunziati, "Automatic detection and recognition of players in soccer videos," in *International Conference on Advances in Visual Information Systems*. Springer, 2007, pp. 105–116.

[4] L. Ballan, M. Bertini, A. D. Bimbo, and W. Nunziati, "Soccer players identification based on visual local features," in *Proceedings of the 6th ACM international conference on Image and video retrieval*, 2007, pp. 258–265.

[5] S. Gerke, K. Muller, and R. Schafer, "Soccer jersey number recognition using convolutional neural networks," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2015, pp. 17–24.

[6] G. Li, S. Xu, X. Liu, L. Li, and C. Wang, "Jersey number recognition with semi-supervised spatial transformer network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 1783–1790.

[7] H. Liu and B. Bhanu, "Pose-guided r-cnn for jersey number recognition in sports," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0.

[8] W.-L. Lu, J.-A. Ting, J. J. Little, and K. P. Murphy, "Learning to track and identify players from broadcast sports videos," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 7, pp. 1704– 1716, 2013.

[9] A. Senocak, T.-H. Oh, J. Kim, and I. So Kweon, "Part-based player identification using deep convolutional representation and multi-scale pooling," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 1732–1739.

[10] S. Messelodi and C. M. Modena, "Scene text recognition and tracking to identify athletes in sport videos," *Multimedia tools and applications*, vol. 63, no. 2, pp. 521–545, 2013.

[11] C.-W. Lu, C.-Y. Lin, C.-Y. Hsu, M.-F. Weng, L.-W. Kang, and H.-Y. M. Liao, "Identification and tracking of players in sport videos," in *Proceedings of the Fifth International Conference on Internet Multimedia Computing and Service*, 2013, pp. 113–116.

[12] M. Sari, H. Dujmi, V. Papi, and N. Roˇ zi, "Player number localization andˇ recognition in soccer video using hsv color space and internal contours," in *The International Conference on Signal and Image Processing (ICSIP 2008)*, 2008.

[13] S. Nag, R. Ramachandra, P. Shivakumara, U. Pal, T. Lu, and M. Kankanhalli, "Crnn based jersey-bib number/text recognition in sports and marathon images," in *2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2019, pp. 1149–1156.

[14] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.

[15] M. Ramanathan, W.-Y. Yau, and E. K. Teoh, "Human body part detection using likelihood score computations," in *2014 IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM)*. IEEE, 2014, pp. 160–166.

[16] D. Deng, H. Liu, X. Li, and D. Cai, "Pixellink: Detecting scene text via instance segmentation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[17] X. Wang and J. Yang, "Marathon athletes number recognition model with compound deep neural network," *Signal, Image and Video Processing*, vol. 14, no. 7, pp. 1379–1386, 2020.

[18] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[19] Z. Tian, W. Huang, T. He, P. He, and Y. Qiao, "Detecting text in natural image with connectionist text proposal network," in *European conference on computer vision*. Springer, 2016, pp. 56–72.

[20] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[21] Y. Baek, B. Lee, D. Han, S. Yun, and H. Lee, "Character region awareness for text detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9365–9374.

[22] J. Baek, G. Kim, J. Lee, S. Park, D. Han, S. Yun, S. J. Oh, and H. Lee, "What is wrong with scene text recognition model comparisons? dataset and model analysis," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 4715–4723.

[23] A. Chan, M. D. Levine, and M. Javan, "Player identification in hockey broadcast videos," *Expert Systems with Applications*, vol. 165, p. 113891, 2020.

[24] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and vision computing*, vol. 22, no. 10, pp. 761–767, 2004.

[25] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[26] S. Gerke, A. Linnemann, and K. Muller, "Soccer player recognition¨ using spatial constellation features and jersey number recognition," *Computer Vision and Image Understanding*, vol. 159, pp. 105–115, 2017.

[27] C. De Vleeschouwer, F. Chen, D. Delannay, C. Parisot, C. Chaudy, E. Martrou, A. Cavallaro *et al.*, "Distributed video acquisition and annotation for sport-event summarization," *NEM summit*, vol. 8, 2008.